European Journal of Political Economy 60 (2019) 101809

Contents lists available at ScienceDirect

European Journal of Political Economy

journal homepage: www.elsevier.com/locate/ejpe

# Corruption, workforce selection and mismatch in the public sector[☆]

Sauro Mocetti, Tommaso Orlando [*]

*Bank of Italy, Structural Economic Analysis Directorate, Economics and Law Division, Italy*

ARTICLE INFO

*JEL classification:*
D73
H83
J45

*Keywords:*
Corruption
Selection
Mismatch
Schooling
Human capital
Public employment

ABSTRACT

We examine the impact of corruption on workforce selection and personnel allocation in the public sector. Using Italian data, we find that the selection of public employees in terms of human capital worsens in comparison to that of their private sector counterparts in areas with higher levels of corruption. Moreover, corruption is associated with educational mismatch in the allocation of human resources and, in particular, with an increase in the rate of under-qualification. These results are robust to several alternative indicators and specifications, including IV estimation using past dependence on public spending and the historical relevance of foreign domination as exogenous sources of variation for current corruption.

## 1. Introduction

Corruption is widely believed to entail large economic and social costs. The economic literature has so far explored several channels through which corruption may affect economic outcomes. Some authors highlight its effects in terms of distortion of private decisions, such as investments (Shleifer and Vishny, 1993; Mauro, 1995) and human capital accumulation (Mo, 2001). Others focus on the activities of the public sector, documenting relationships between corruption and inefficiencies in the composition of government expenditure (Mauro, 1998), lower productivity of public investments (Del Monte and Papagni, 2001) and higher shares of goods and services procured by the public administration on non-competitive markets (Hessami, 2014).

In this paper we analyze the impact of corruption on workforce selection and personnel allocation in the public sector. More specifically, we address two issues: first, we examine whether corruption affects the *selection into the public sector* of individuals with different levels of human capital; second, we examine the relationship between corruption and the *allocation within the public sector* of differently educated individuals to jobs with different skill content. Poorer recruitment and misallocation of human resources within public agencies might have significant and long-lasting consequences on the quality of the administration's activity and on the effectiveness of services provided by the public sector. Nevertheless the impact of corruption through these channels has so far

☆ This paper circulated in working paper version as 'Corruption and personnel selection and allocation in the public sector', *Bank of Italy Occasional Papers n. 402*. We thank Audinga Baltrunaite, Federico Cingano, Silvia Giacomelli, Laura Ogliari, Giuliana Palumbo, Paolo Sestito, Luigi Federico Signorini and participants at the conference on "Corruption, Tax Evasion and Institutions" (Riga, 2017), at the SIEP conference (Catania, 2017), at the SIE conference (Rende, 2017) and at Bank of Italy seminars for useful comments. The views expressed in this paper should be referred only to the authors and not to the Institution with which they are affiliated.

* Corresponding author. via Nazionale 91, 00184, Rome. Italy.

*E-mail address:* tommaso.orlando@bancaditalia.it (T. Orlando).

https://doi.org/10.1016/j.ejpoleco.2019.07.007
Received 28 August 2018; Received in revised form 17 July 2019; Accepted 21 July 2019
Available online 24 July 2019

remained surprisingly uninvestigated. Moreover, corruption might also induce rent-seeking activities, involving unproductive use of resources and causing social losses (Aidt, 2016).

The empirical analysis is based on two complementary data sources containing information on Italian public and private employees and exploits several measures of corruption. We examine whether areas characterized by higher values of our corruption indicators show peculiar patterns of skill-based selection into and allocation within the public sector. Although we rely on cross-sectional variation, the empirical strategy by which we study workforce selection is made robust to potential omitted-variables bias by the inclusion of a rich set of geographical and profession-related fixed effects. The impact of corruption on personnel allocation is addressed through a specification that mirrors a difference-in-differences approach, in which the treatment is represented by the intensity of corruption at the local level, and exposure to the treatment is determined by whether workers are employed in the public or private sector. Moreover, to address the obvious reverse causality issue – the possibility that corruption itself is the consequence of poor selection and allocation of human resources by the public administration – we instrument corruption with two variables that are positively correlated with corruption, but which predate the hiring of current public employees: past dependence on public spending and the historical relevance of foreign domination.

We find that public employees are, on average, more educated and obtained higher grades at school than their professional counterparts in the private sector. However, in areas with higher corruption the relationship between educational attainments and the likelihood of working in the public sector is substantially weaker. The negative impact of corruption is concentrated among those with higher skill content jobs, such as managers and highly skilled professionals. As for the allocation process, we find that a higher level of corruption is associated with an increase, relative to the private sector, in the likelihood of mismatch between individual educational attainments and the skill content of the job one is assigned to. This mismatch comes mostly in the form of under-education – individuals being assigned to jobs that are, on average, undertaken by more qualified personnel – rather than over-education. We also show that under-education is not only a "mechanical" consequence of poorer selection processes or of inflation in the number of managerial positions. Finally, we show that corruption also leads to a relatively lower effort by public employees, as measured by hours worked. As a whole, our findings point to corruption inducing a lower quality and a reduced efficiency of the public sector.

The literature has already partially dealt with the relationship between corruption and occupational choices. Murphy et al. (1991) and Acemoglu and Verdier (1998) argue that corruption magnifies rewards to rent-seeking activities and subtracts valuable human resources to entrepreneurship. Concerning selection into the public sector, experimental evidence suggests that more corrupt environments encourage entry by dishonest people into the public sector: Banerjee et al. (2015) and Hanna and Wang (2017) find negative self-selection into the Indian public administration, while Barfort et al. (2017) find positive self-selection into the Danish public sector. Colonnelli et al. (2018) find that political patronage is a key determinant of employment in the public sector.

Our paper innovates upon the existing literature along several directions. First, the economic impact of corruption has typically been investigated using cross-country evidence (at a single point in time). However, cross-sectional relationships might be severely biased, as corruption and the other variables of interest are likely to have common correlates that cannot all be credibly controlled for.[1] To address this issue, some papers introduce country fixed effects by exploiting panel data. However, the reliability of those estimates clearly depends on the longitudinal (within-country) variation of these factors that, in the case of a persistent and structural phenomenon like corruption, is arguably low. Moreover, panel data alone do not fully address endogeneity concerns, as a variation in corruption and in the outcome variable might reflect common, country-specific shocks. To tackle these difficulties, we exploit a different identification strategy that hinges on the differential impact of corruption among individuals living in the same area, while controlling through area fixed effects for any other potential omitted variable correlated with corruption.

Second, the measurement of corruption itself may be problematic either from a cross-sectional or a longitudinal point of view. Indeed, one may question the capacity of international surveys to capture the intensity of corruption equally well in all countries, due to differences in culture and social norms or to other perception biases. Similarly, official data on reported crimes might not be comparable across countries due to differences in laws or in the availability of harmonized crime statistics. The extent of these measurement issues can also vary over time. However, our analysis is based on various measures of corruption within a single country, thus exploiting (sub-national) territorial variability while using homogeneous and comparable indicators.

Third, previous studies on the relationship between corruption and workforce sorting have used experimental evidence and focused on whether workers' personal propensity to dishonesty makes them more likely to self-select into the public sector. In contrast, we rely on hard data and drive the attention towards the impact of environmental levels of corruption on sorting based on human capital and personal allocation.[2] Personal attitudes towards unlawful behavior are undoubtedly of the foremost importance in determining the quality and impartiality of public services. However, poor human capital endowments and occupational mismatch might also hinder the effectiveness of economic decisions by public agencies on a number of relevant dimensions, such as the level and the composition of public expenditure, the effectiveness of public investments and the quality of services provided to households and firms.

[1] Stated differently, less corrupt societies appear to perform well in almost any dimension, and the risk of bias due to an omitted variable (e.g., of institutional or cultural nature) is large.

[2] It is worth noting that selection and allocation processes are both important, even if previous studies have mainly directed their attention towards workforce sorting only. Indeed, on the one hand, the same group of individuals can produce substantially different results if they are badly matched to jobs requiring different educational qualifications. On the other hand, bad allocation processes and misaligned career rewards might discourage the most skilled individuals from applying for a public job in the first place.

Finally, even if we cannot observe individual performance, we are able to examine the impact of corruption on hours worked, a proxy of effort exerted by employees. We find that corruption is associated with lower levels of effort in the public sector, relative to the private sector.

The rest of the paper is organized as follows. Section 2 describes the data sources and the main variables of our analysis, including the construction of corruption indicators. Section 3 details the empirical strategy. Section 4 presents our main findings and some robustness checks. Section 5 concludes.

## 2. Data and variables

### 2.1. Individual information on occupation and schooling

Individual data on employment characteristics and observed measures of human capital are drawn from two sources. The main one is the Italian Labour Force Survey (LFS). The survey is carried out by the National Institute of Statistics (Istat) and its main aim is to provide official statistics on the employed and unemployed population in Italy. We pool the LFS waves from 2004 to 2010 and we restrict the analysis to non-manual employees (i.e., ISCO major groups 1 to 5).[3] LFS does not provide a clean distinction between the public and the private sector and, therefore, we identify as public employees all those employed in the following three NACE 2-digits groups: public administration, education and healthcare. We also know the professional qualification of each employee, as indicated by the ISCO occupational classification at 3 digits, and their education level (in particular, the years of schooling corresponding to their highest educational attainment). This information allows us to construct an individual indicator of under-education (over-education): employees are under-educated (over-educated) if their educational attainments (measured by completed years of schooling) lies below the 25th percentile (above the 75th percentile) of the years of schooling distribution within their professional category.

Beyond the overall picture, we also provide evidence on the subgroup of managers and professionals (ISCO major groups 1 and 2), i.e. those employees who are at the top of the occupational hierarchy and who are responsible for controlling or managing an organization or staff teams. This focus is motivated by the fact that managers can single-handedly shape the activity of the public agency. We also observe hours worked in the week preceding the interview and, among socio-demographic characteristics, age, gender and, most importantly for our goal, the local labor market (henceforth LLM) in which workers reside.[4] This is our geographic unit of analysis. This geographic attribute also captures local economic and social conditions that might impact on the likelihood of working in the public sector and on the quality of the match between individual education and job skill content.

The selection process is also investigated through the use of a second data source, the Survey on Household Income and Wealth (SHIW). The survey is carried out by the Bank of Italy and contains information on the socio-economic conditions of a representative sample of the Italian population.[5] We pool the (bi-annual) SHIW waves from 2000 to 2014 and we restrict the analysis to household heads who are non-manual employees, as done with the LFS. The size of the SHIW sample is much smaller than that of the LFS and details on occupation are definitely poorer. However, unlike the LFS, the SHIW allows a clean distinction between public and private sector. More importantly, the SHIW can be used to complement the LFS analysis with a further dimension of individual human capital: the final grade associated with the individual's highest educational attainment. Among socio-demographic characteristics, we include – as before – age, gender and the LLM where the individual resides.

Descriptive statistics for the LFS and SHIW samples are reported in Table 1. Consistently with evidence from previous studies (see for instance Giorgiantonio et al., 2016, and Rizzica, 2016), in both samples public sector employees are, on average, older, include a larger share of women and possess relatively richer endowments of human capital, both in terms of educational attainments and grades obtained at school. Moreover, the extent of average mismatches (especially under-education) is similar in the private and the public sector. Finally, public employees work on average fewer hours per week than private employees.

### 2.2. Measures of corruption

Measuring corruption is admittedly a challenging task: as all illegal activities, corruption is mostly unobservable and, therefore, difficult to quantify. There are several possible definitions of corruption and different approaches to its measurement. These approaches can be classified along two dimensions. First, they may be either *subjective* or *objective*, depending on whether they rely on – respectively – survey or 'hard' data. Second, they may be either *direct* or *indirect*, according to whether they quantify – respectively – corruption itself or other variables deemed to be 'naturally correlated' to corruption, and perhaps easier to observe.

The four approaches that emerge from combinations along these two dimensions have all, to some extent, been used to quantify corruption. Subjective and direct indicators include, for example, Transparency International's Corruption Perceptions Index and the Global Corruption Barometer, as well as the European Quality of Government Index (see Charron et al., 2014). The second approach relies on subjective but indirect indicators of corruption: for instance, distrust towards local or national governments might at least

[3] The elaboration on Istat data for this work have been carried out at the Istat Elementary Data Analysis Laboratory (ADELE) in accordance with the legislation on statistical confidentiality and personal data protection. Reported results and their interpretation are to be considered sole responsibility of the authors and do not by any means represent official statistics or involve Istat in any other way. All analyses reported hereunder do not make use of sample weights.

[4] LLMs are 686 geographic units composed of contiguous municipalities and delimited on the basis of daily commuting patterns. Therefore, a LLM represents the area in which most individuals both reside and work and it is the most proper geographical unit for labor market analysis.

[5] See Brandolini and Cannari (1994) for more details on the survey.

**Table 1**
Descriptive statistics.

| | *Full sample* | | *Public sector employees* | |
|---|---|---|---|---|
| | Mean | St. dev. | Mean | St. dev. |
| **LFS data** | | | | |
| Female | 0.561 | 0.496 | 0.656 | 0.475 |
| Young (<35 y.o.) | 0.293 | 0.455 | 0.149 | 0.356 |
| Years of schooling | 12.535 | 3.337 | 13.537 | 3.467 |
| Under-education | 0.126 | 0.332 | 0.124 | 0.329 |
| Over-education | 0.090 | 0.286 | 0.103 | 0.304 |
| Hours worked | 34.859 | 10.348 | 31.883 | 10.012 |
| Observations | 753,048 | | 301,120 | |
| **SHIW data** | | | | |
| Female | 0.365 | 0.482 | 0.382 | 0.486 |
| Young (<35 y.o.) | 0.110 | 0.313 | 0.073 | 0.260 |
| Years of schooling | 13.086 | 3.256 | 13.433 | 3.481 |
| Grades obtained at school | 0.824 | 0.136 | 0.837 | 0.138 |
| Observations | 11,511 | | 5905 | |

Years of schooling are those corresponding to the highest educational attainment. Grades are the final grades (normalized with respect to the maximum obtainable grade) obtained at the highest educational attainment (they are available only for individuals with at least a diploma). Under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution within his/her profession (based on the ISCO 3-digit classification). The statistics on hours worked refer to the week prior to the survey and exclude individuals that worked 0 h.

partially reflect their being perceived as corrupt entities by citizens. An example of objective though indirect measures of corruption is "missing expenditure" (Olken, 2009) i.e., the difference between public expenditure for infrastructures and their value when realized, which is arguably an observable consequence of corruption. The fourth approach relies on objective and direct measures of corruption such as officially recorded corruptive crimes, police investigations, or analogous evidence.[6]

Each of these approaches has advantages and drawbacks. Perception-based indicators (either direct or indirect) have been widely used, as they are available for a large set of countries and allow to exploit cross-country variation in corruption to examine the latter's relationship with other economic outcomes.[7] However, the effectiveness of these indicators has been questioned. First, there are significant differences in cultural traits, social norms and laws across countries, so that citizens of one polity may find certain practices more acceptable than citizens of another, thus leading to different reported perceptions of the extent of corruption. Second, the reliability of survey information has also been questioned, as respondents might not report direct experiences but be influenced by what is publicized in the media (Rizzica and Tonello, 2015). The objective indirect approach is also intriguing, but missing expenditure – like any other variable measured as a "residual" – is not necessarily attributable to corruption. For example, the effectiveness of public spending in infrastructures might also reflect the efficiency of the local construction industry, unobserved characteristics of the territory or other random elements, thus confounding the interpretation of the computed indicator. Finally, the object direct approach, beyond poor cross-country comparability due to differences in laws and in the organization of the judicial system, might suffer from reporting bias. If crime episodes are collected by police forces or courts, variations in their number might reflect not only the intensity of the criminal activity, but also the efficiency of these institutions and/or their interest in prosecuting that particular type of offence.

In this paper we adopt two measures of corruption. The first is based on corruption-related (alleged) crimes as reported by police forces, net of local judicial efficiency. Data on reported crimes are extracted from the SDI database,[8] and are observed at the municipality level for the period 2004–2011. In particular, we restrict the analysis to crimes intimately linked to corruptive practices: corruption proper (bribery), graft and malfeasance.[9] These unlawful behaviors all result into additional payoffs accruing to the public employee at the detriment of one or more private agents: the role of such agents may range from being an active part in the enactment of the criminal deed (as in bribery) to being the victims of the civil servant's prevarication (as in graft). These raw figures on crimes are normalized with respect to total employment at the local level (a proxy for the level of economic

[6] Olken and Barron (2009) designed a study in which surveyors accompanied Indonesian truck drivers on their trips in order to collect direct observations on illegal payments to police, soldiers, and weigh station attendants. Ferraz and Finan (2011) and Brollo et al. (2013) use data on a program of random audits on local governments, with detailed reports on corruption charges. For Italy, Del Monte and Papagni (2001, 2007) and Barone and Mocetti (2014) use official statistics on reported crimes against the public administration.

[7] See, among the others, Mauro (1995), Knack and Keefer (1995), La Porta et al. (1999), Fisman and Gatti (2012) and Fisman and Miguel (2007).

[8] SDI (*Sistema Di Indagine*) is managed by the Ministry of Interior and collects data on all investigations managed by the three main Italian police forces (*Arma dei Carabinieri*, *Polizia di Stato* and *Guardia di Finanza*). Information available to us consists of the number of (alleged) criminal offences by type of offence, municipality and year.

[9] Crimes perpetrated by public officials are regulated by the Italian criminal law (*Codice Penale*, articles 314–323, 479–481 and 493). Acknowledging oversimplification, *corruption proper* ('*corruzione*' in Italian) takes place when the public official accepts a bribe from a private counterparty in exchange for the enactment of or the abstention from certain behaviors. *Graft* ('*concussione*') refers to the situation in which the payment is imposed by the civil servant to the private party. Here *malfeasance* ('*abuso d'ufficio*') generically defines behaviors enacted by the public employee aiming at earning unlawful benefits: resource embezzlement ('*peculato*') and document forgery ('*falsità ideologica*'), when perpetrated by public officials or by other providers of public services, may be seen as special cases of malfeasance.

**Table 2**
Corruption: principal component analysis.

| | 1st component | 2nd component | 3rd component | 4th component |
|---|---|---|---|---|
| Eigenvalue | 2.573 | 0.883 | 0.330 | 0.214 |
| Proportion | 0.643 | 0.221 | 0.082 | 0.054 |
| Cumulative | 0.643 | 0.864 | 0.946 | 1.000 |
| Coefficient 1st component | $C^1$ | Trust | GP | CPI |
| | 0.365 | 0.482 | 0.382 | 0.486 |

transactions). Our corruption measure is computed at the local labor market (LLM) level and is averaged over the period of observation[10]. Clearly, police reports of corruption-related crimes may reflect, on top of the underlying criminal phenomenon, the intensity with which local judiciary authorities prosecute these crimes. This consideration is especially relevant in Italy, where – despite uniform national regulatory frameworks – judicial efficiency varies widely across courts. To address the potential reporting bias, we partial out any effect of local judicial efficiency on crime rates. Namely, we run a regression of our corruption indicator on local judicial efficiency (measured by the lengths of criminal trials in the relevant court), and we construct our final indicators by taking the residuals thus obtained. This procedures yields a measure of corruption incidence net of local judicial efficiency ($C^1$ henceforth).[11]

The second measure is a synthetic indicator that combines information drawn from different approaches. Namely, we collect four different variables, each echoing one of the four measurement methods mentioned above (though, for reasons of data availability, they are measured at different geographical levels). The first variable is a subjective assessment of the level of corruption (*CPI*). Data are drawn from a large European survey (EQI) aimed at measuring the quality of governance within the European Union, and they are available at the regional level.[12] The second variable echoes the subjective and indirect approach to the measurement of corruption. We exploit a survey managed by Istat (the so-called "Multiscopo") asking a large set of questions to citizens on various aspects of life, including trust towards local government and other institutions (*TRUST*). Based on a rich literature on the detrimental effect of perceived corruption on the trust expressed by people towards local institutions (e.g., Uslaner, 2004; Clausen et al., 2011; Uslaner, 2004 and Clausen et al., 2011), we take distrust towards local government as an indicator of corruption. These figures are available at the regional level with a further distinction between small municipalities, intermediate municipalities and larger metropolitan areas.[13] The third variable belongs to the group of objective and indirect measures of corruption. Golden and Picci (2005) compute a measure of corruption for Italy based on the difference between the value of public infrastructure and cumulated public expenditure in public works (*GP*). These figures are available at the regional level. Our last variable is reported crime adjusted for judicial efficiency, the aforementioned $C^1$. We then rely on a principal component analysis to extract information from these four variables. The first principal component explains about 64 percent of the total variance of the underlying variables and it is positively associated, as expected, with each one of the input variables (Table 2). We call this synthetic indicator $C^2$.

$C^1$ and $C^2$ both have some strong and weak points. On the one hand, $C^2$ might better capture a multidimensional and not easily observed phenomenon such as corruption. Moreover, the large fraction of variance explained by the first component suggests that the four indicators largely overlap, which is supportive of the measure's rich informational content. On the other hand, $C^1$ is easier to interpret in economic terms and less subject to arbitrary choices. In addition, while both measures vary at the LLM level, $C^1$ exclusively reflects variations in observed corruption at that very level, while $C^2$ partly reflects indicators that vary at a much broader level. For these reasons, $C^1$ is our preferred measure of corruption, though we provide evidence using both indicators throughout the paper.[14]

Summary statistics of the two indicators are reported in Table 3. In order to guarantee comparability between different indicators of corruption, both $C^1$ and $C^2$ are standardized. The two variables display considerable variability across LLMs. A graphical representation of the territorial differences in corruption intensity is reported in Fig. 1: both indicators show that corruption is more widespread in Southern Italy, with the North-South divide being more visually evident when $C^2$ is used, as the latter by construction amplifies regional differences. However, in both cases there is also significant variability within each of the three statistical macro-areas (North, Center, and South).

[10] We do not exploit the time dimension since corruption is a persistent phenomenon and does not show sufficient longitudinal variation.

[11] According to our findings, a variation of one standard deviation of the length of criminal trials is associated to a 0.14 standard deviations increase in the reported corruption rate, suggesting that the latter largely reflects the intensity of criminal activity at the local level, and is only marginally affected by judicial efficiency.

[12] Italy's 110 provinces are grouped into 20 regions. More information on the data can be found in Charron et al. (2014).

[13] Since not every region includes a large metropolitan area, this variable takes on 51 distinct values, rather than 60 as one might expect.

[14] Our main results are qualitatively confirmed even if we use raw figures for reported corruption (i.e., without correcting for judicial efficiency) or each indicator used in the principal component analysis separately. Results are available upon request.

**Table 3**
Corruption: descriptive statistics.

| | Mean | S.D. | 10th | 25th | 50th | 75th | 90th |
|---|---|---|---|---|---|---|---|
| Crime rate: $C^1$ | 0.000 | 1.000 | −0.570 | −0.442 | −0.254 | 0.141 | 0.765 |
| Principal component: $C^2$ | 0.000 | 1.000 | −1.299 | −0.739 | 0.133 | 0.856 | 1.225 |

Corruption indicators are standardized at the LLM level.

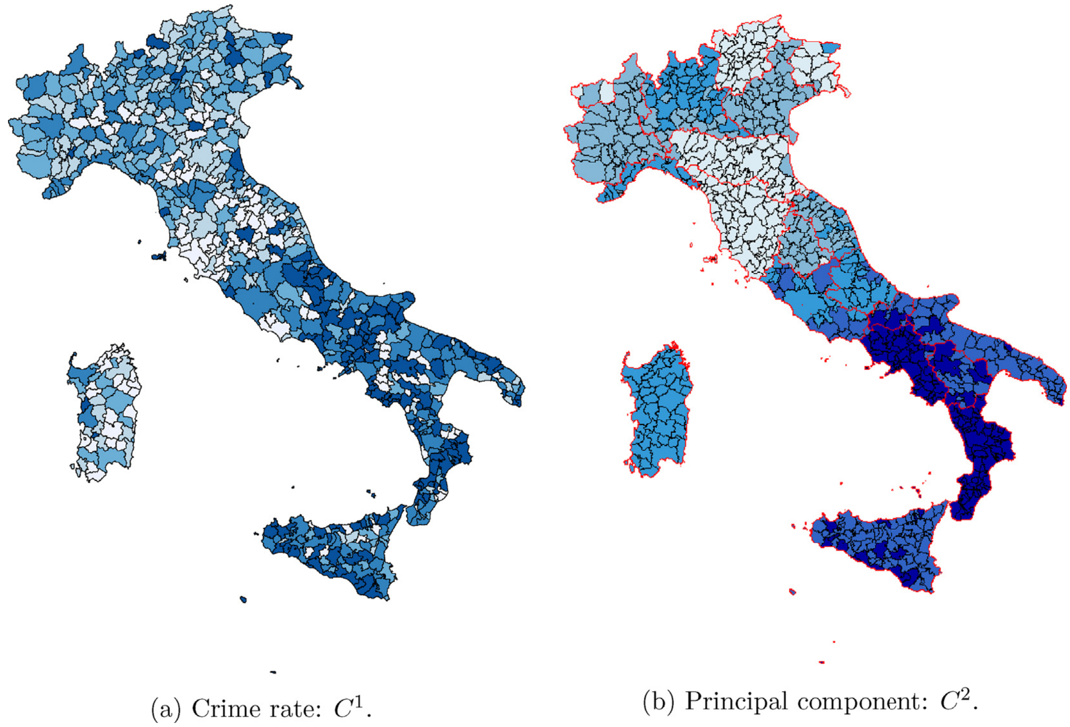


**Fig. 1.** A map of corruption at the LLM level (black borders). Panel (b) also reports regional boundaries (red borders) to highlight the impact of including variables measured at the regional level. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

### *2.3. Descriptive evidence*

Corruption and human capital endowments are positively correlated at the LLM level, as shown in Fig. 2. This apparently surprising fact is mainly due to underlying factors being related to both variables. For example, corruption is positively associated with the size of the public sector (Table 4), either because large public agencies offer better chances to corrupts or because corruption may hinder the development of more market-oriented activities. Sector composition, however, also affects the incentives to invest in human capital: as it has been widely documented, the public sector tends to attract the most educated workers (Cowley and Smith, 2014; Rizzica, 2016). Moreover, corruption is more widespread in less economically developed LLMs, as measured by the value added per capita, and by poorer labor market opportunities. The latter, however, might also affect human capital investments reducing *ceteris paribus* the opportunity cost of studying. Indeed, if one controls for the above-mentioned variables, the correlation between corruption and education disappears. This also suggests that, when attempting to identify a clean effect of corruption on other socio-economic outcomes, we face the challenging task of having to avoid spurious correlation driven by unobserved omitted variables.

Table 4 provides comparisons between employee characteristics in the public and private sectors across areas with different levels of corruption. Having established that the share of employees in the public sector is larger in LLMs where the intensity of corruption is higher, we find that the share of managers among public employees in those areas is also slightly larger. In terms of human capital, as measured by the years of schooling, the (positive) gap between public and private sector is narrower in more corrupt LLMs. Furthermore, the (negative) difference between under-education in the public and private sector is also reduced (in absolute value) when one moves from less to more corrupt areas. Finally, the (negative) gap in total hours worked between public and private sector grows larger in more corrupt areas, especially for managers.

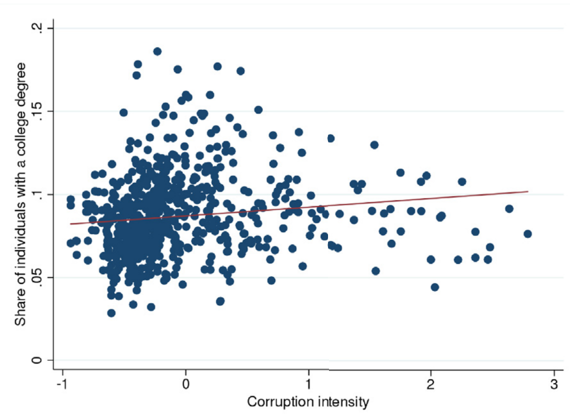


**Fig. 2.** Corruption and education across LLMs. Corruption intensity is measured by $C^1$; the share of population with a college degree is drawn from the 2001 Census.

**Table 4**
Comparison between low- and high-corruption areas.

| | *Low-corruption LLMs* Mean | *High-corruption LLMs* | Difference |
|---|---|---|---|
| Public employees | 0.372 | 0.429 | 0.06*** |
| Managers in public sector | 0.266 | 0.311 | 0.04*** |
| Observations | 382,607 | 370,441 | |
| **All employees** | | | |
| Diff. years of schooling | 0.463 | 0.357 | −0.11*** |
| Diff. under-education | −0.050 | −0.028 | 0.02*** |
| Diff. hours worked | −2.353 | −2.632 | −0.28*** |
| Observations | 382,607 | 370,441 | |
| **Managers** | | | |
| Diff. years of schooling | 1.429 | 1.280 | −0.15*** |
| Diff. under-education | −0.180 | −0.169 | 0.01*** |
| Diff. hours worked | −2.517 | −3.877 | −1.36*** |
| Observations | 61,386 | 73,744 | |

LLMs are classified according to whether their value of $C^1$ exceeds the national median. Variables labeled 'Diff.' measure differences between averages in the public and private sector. Years of schooling are those corresponding to the highest educational attainment. Under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution within his/her profession (based on the ISCO 3-digit classification). The statistics on hours worked refer to the week prior to the survey and exclude individuals that worked 0 h. Tests for differences in means allow for unequal variance in the two sub-samples.

## 3. Empirical strategy

### *3.1. Selection of workers into the public sector*

The first phenomenon we address is the potential distortionary effect of corruption on the relevance of educational attainments as predictors of the individual likelihood of working in the public sector. To this end, we estimate[15] the following linear probability model:

$$Y_i = \alpha + \beta S_i + \delta \left( S_i \cdot C_{LLM(i)} \right) + \gamma' X_i + \rho_{LLM(i),p(i)} + \varepsilon_i \tag{1}$$

where $Y_i$ is a binary indicator of the occupational status of individual $i$, taking on the value of 1 if $i$ is a public employee and the value of 0 if $i$ is employed in the private sector; $S_i$ is a measure of $i$'s human capital endowment, e.g. the completed years of schooling; $C_{LLM(i)}$ is one of the two measures of the incidence of corruption in the LLM in which individual $i$ resides; $X_i$ is a vector of individual controls such as gender and age: these are included as the likelihood of having joined the public sector may be affected by gender- or cohort-specific factors; finally, the term $\rho_{LLM(i),p(i)}$ is a group indicator, obtained by combining $i$'s LLM and occupation (ISCO groups from 1 to 5), capturing unobserved factors at that level. With the latter set of fixed effects we control for any potential (time-invariant) omitted variable at the LLM level affecting the choice between the public and private sector plus, within each LLM, for any other variables affecting the same choice in occupation-specific labor markets (e.g., managers or clerical support workers).

[15] Estimations are performed using the methodology and software presented in Correia (2017).

Thus, our coefficient of interest $\delta$ captures how the impact of schooling on the likelihood of belonging to a certain professional area $p(i)$ in the public sector (rather than the same occupation in the private sector) varies across LLMs characterized by different corruption intensity. We might expect $\delta < 0$, indicating that corruption weakens the positive relationship between education and the likelihood to join the public sector. As $S_i$ is predetermined with respect to the observed occupation $Y_i$, our specification represents a stylized occupational choice model.

We expect the impact of corruption to differ across individuals with different endowments of human capital for several reasons. First, schooling is usually associated with other positive social behaviors such as, among many, an increased likelihood to vote, a reduced likelihood of participating in crime and of being arrested, and an increased tendency to behave legally (Oreopoulos and Salvanes, 2011).[16] Therefore we might expect that highly educated individuals have a lower propensity to self-select themselves in the public sector (instead of the private sector) in areas where administrations are more corrupt. Second, bribes' returns are presumably higher for less educated workers, who have worse outside opportunities and, therefore, lower opportunity costs (Lochner, 2004). Finally, it is also likely that more corrupted public administrations will select, and later remunerate individuals more in terms of soft skills, relational capital or craftiness, rather than according to abilities proxied by educational attainments.

### 3.2. Allocation of workers within the public sector

Besides affecting, through self-selection and screening, the composition of the available public workforce, corruption may have an impact on how efficiently these human resources are assigned to different jobs and tasks. In particular, we posit that, for each job, there exists an optimal level of skills or human capital, i.e., the level of an individual which is "just right" for that job. We proxy such level by the nation-level average education of workers in that position. We subsequently test whether corruption shifts the allocation of human resources away from the right matching and, if that happens, whether this prevalently takes the form of under- or over-education, i.e., of employees having respectively a much lower or higher skill level than that required on average by the jobs they are assigned to. In order to quantify this effect, we estimate the following linear probability model:

$$M_i = \alpha + \beta \left(Y_i \cdot C_{LLM(i)}\right) + \gamma' X_i + \rho_{LLM(i)} + \varphi_{s(i)} + \varepsilon_i \tag{2}$$

where the dependent variable $M_i$ is a binary indicator for the presence of some form of skills mismatch (under-education or over-education) for individual *i*. Specifically, an individual is considered to be under-educated if her schooling level falls below the 25th percentile of the distribution of schooling within her profession (defined in terms of the ISCO 3-digit classification) and, conversely, over-educated if her schooling level exceeds the 75th percentile of that distribution. $Y_i$ denotes whether $i$ is employed in the public or in the private sector. LLM-fixed effects ($\rho_{LLM(i)}$) and sector-fixed effects ($\varphi_{s(i)}$) capture local or industry-specific variables that might be correlated with mismatch.[17] Our coefficient of interest is $\beta$, which captures how the impact of working in the public sector on the likelihood to be mismatched varies across LLMs characterized by different corruption intensity. We might expect $\beta > 0$, as corruption is more likely to generate incentives for public managers, rather than for private ones, to deviate from the most efficient personnel allocation.

Our specification recalls a difference-in-differences approach, in the same spirit of the methodology pioneered by Rajan and Zingales (1998).[18] Namely, $C_{LLM(i)}$ represents the treatment and $Y_i$ assigns units to groups differently exposed to the treatment. We assume that the public sector is more exposed to corruption with respect to the private sector. Indeed, corruption necessarily implies the involvement of a public official, while private sector workers might be more of less exposed to corruption, depending on the extent of the interaction between their industry and the public sector. In particular, we assume that the manufacturing sector, being hardly dependent on public spending and more exposed to international competition, is the least affected by corruption. In contrast, other private economic activities (e.g., public utilities, construction sector, highly regulated industries, etc.) are more exposed to corruption as they are more dependent upon public spending and, therefore, more involved in some kind of interaction with public officials.

### 3.3. Identification assumptions

When we examine the impact of corruption on economic outcomes exploiting cross-sectional variation, we should take account of two potential identification threats.

First, unobserved heterogeneity at the local level (e.g., social norms, level of economic development, etc.) might be related to corruption, to the patterns of individual human capital accumulation, and to our dependent variables. These omitted variables are likely to bias OLS estimates. In order to address this issue, we saturate the model with a rich set of controls. On the one hand, we include LLM-fixed effects that should capture the impact of all potential omitted factors varying at that level. On the other hand, we also control for within-LLM variation. For example, in equation (1) we also control for LLM and professional area *interacted* fixed effects, so taking into account schooling-biased effects of corruption across homogeneous professional areas within each LLM; in equation (2) we also include, in a robustness check, the average level of schooling within each LLM $\times$ sector cell, so controlling for LLM-specific sorting across sectors as a function of education. Finally, in another exercise, we also control for other variables

[16] With specific reference to the Italian case, education is found to be positive associated to individual tax morale (Barone and Mocetti, 2011).

[17] Sectors are observed at the NACE 2 digit level.

[18] A similar approach can be found also in Madrian (1994).

**Table 5**
Selection: the impact of schooling (LFS).

| Dependent variable: | Employed in the public sector | | | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Years of schooling | 0.016*** | 0.025*** | 0.015*** | 0.024*** |
| | (0.001) | (0.001) | (0.001) | (0.001) |
| Years of schooling × $C^1$ | −0.008*** | −0.013*** | | |
| | (0.001) | (0.002) | | |
| Years of schooling × $C^2$ | | | −0.006*** | −0.009*** |
| | | | (0.001) | (0.001) |
| LLM × Prof. area FE | X | X | X | X |
| Observations | 396,764 | 99,644 | 396,764 | 99,644 |
| Adj. R-squared | 0.27 | 0.22 | 0.27 | 0.22 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, *** $p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: $C^1$ – i.e., reported crimes net of judicial efficiency – and $C^2$ – i.e., the principal component of *CPI*, *TRUST*, *GP* and $C^1$. Other controls include fixed effects for gender, cohort, and survey wave.

(plausibly correlated with corruption) included with an interaction term, on the grounds that they might have a differential effect similar to that produced by corruption itself.

Second, we might suspect the presence of reverse causality, as one may argue that skill-biased recruitment and human resources management processes in the public sector could affect the intensity of corruption. To address this problem, we exploit variation in corruption intensity at the local level that is attributable to factors associated with corruption but predating the hiring of the current public employees. Namely, we use two instrumental variables. The first is the local dependence of the private sector on public demand and, therefore, on public spending, measured using the 1971 Census.[19] The idea is that where this dependence is higher, the economic rents associated with the discretionary power of public officials, as well as the incentive for entrepreneurs to influence public conduct, are also higher. This, in turn, may have increased the likelihood of corruptive practices being established. The second instrument is inspired by a large body of literature that investigates how history (and historical institutions) may still influence existing institutions and current social behaviors (e.g., Acemoglu and Robinson, 2012). Using the information reconstructed by Di Liberto and Sideri (2015), we build a measure of the historical duration of foreign domination at the local level.[20] The relevance of this second instrument, which we verify statistically, relies on the consideration that informal institutions and behaviors may have evolved along different lines depending on whether an area was controlled by local or foreign rulers (see, in particular, Di Liberto and Sideri, 2015, and Petrarca and Ricciuti, 2013, for their connection to our topic in the Italian case).

Finally, it is worth noting that, as shown by Moulton (1990), in a regression performed on micro units and including aggregated (in our case LLM-level) variables, OLS standard errors will be underestimated. To address this issue, we cluster standard errors at the LLM level in both equations (1) and (2).[21]

## 4. Results

### 4.1. The impact of corruption on workforce selection

Table 5 reports the results of the estimation of model (1) for our two main measures of corruption, $C^1$ and $C^2$. The sample, drawn from the LFS, include all the employees of the public and manufacturing sectors engaged in non-manual activities. Individual human capital is measured by years of schooling. Higher educational attainments are, as expected, positively associated with the likelihood of having joined the public sector. One additional year of schooling increases the probability of being a public employee by around 1.6 percentage points; the impact is higher among managers and professionals (2.5 percentage points). More interestingly, corruption reduces the role of education as a predictor of being a public employee. According to our results, moving from a LLM at the 10th percentile of $C^1$ to one at the 90th percentile (i.e., from a low-corruption to a high-corruption LLM) the impact of one additional year of schooling decreases from 2.1 to 1.0 percentage points. The detrimental effect of corruption is larger for managerial and professional occupations, for which the same exercise would lead to a decrease in the impact of education from 3.2 to 1.5 percentage points. The last two columns of Table 5 replicates the analysis using $C^2$ as an approximation of corruption intensity at the local level: results are qualitatively similar.

[19] Dependence on the demand of the public sector at the local level is computed in two steps. First, using the input-output matrix, we compute the dependence on the public demand for each sector of economic activity. Second, we translate these figures at the local level using the historical sector composition of the local economy (i.e., the distribution of employees across sectors at the local level as recorded by the 1971 Census).

[20] More specifically, our instrument equals zero for LLMs that were never under the rule of a foreign power, where *foreign* applies to polities not formally based within the current Italian boundaries. Otherwise, the instrument is the logarithm of the total years of foreign domination.

[21] In the Appendix we also replicate our main results clustering at a wider spatial level (see Table A1). However, standard errors at the province level are fairly similar to those computed at the LLM level: this is not surprising, as in Italy geographical mobility is low and LLM are, by construction, self-contained labor markets.

**Table 6**
Selection: the impact of schooling (SHIW).

| Dependent variable: | Employed in the public sector | | | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Years of schooling | 0.005*** | 0.006*** | 0.005*** | 0.006*** |
| | (0.002) | (0.002) | (0.001) | (0.001) |
| Years of schooling × $C^1$ | −0.003 | −0.004* | | |
| | (0.003) | (0.002) | | |
| Years of schooling × $C^2$ | | | −0.003** | −0.003** |
| | | | (0.002) | (0.001) |
| LLM × Prof. area FE | X | X | X | X |
| Observations | 4939 | 2419 | 4939 | 2419 |
| Adj. R-squared | 0.20 | 0.19 | 0.20 | 0.19 |

Standard errors are clustered at the LLM level (*p < 0.1, ** p < 0.05, ***p < 0.01). The sample includes non-manual employees with at least secondary education, drawn from the SHIW. The dependent variable is equal to 1 for public sector employees and to 0 for industrial sector employees. Years of schooling are those corresponding to the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: $C^1$ – i.e., reported crimes net of judicial efficiency – and $C^2$ – i.e., the principal component of *CPI*, *TRUST*, *GP* and $C^1$. Other controls include fixed effects for gender, cohort, and survey wave.

**Table 7**
Selection: the impact of grades (SHIW).

| Dependent variable: | Employed in the public sector | | | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Grades at school | −0.099** | 0.018 | −0.098*** | 0.015 |
| | (0.046) | (0.040) | (0.045) | (0.039) |
| Grades at school × $C^1$ | 0.037 | −0.110* | | |
| | (0.052) | (0.059) | | |
| Grades at school × $C^2$ | | | −0.006 | −0.077*** |
| | | | (0.028) | (0.024) |
| LLM × Prof. area FE | X | X | X | X |
| Observations | 4926 | 2414 | 4926 | 2414 |
| Adj. R-squared | 0.21 | 0.19 | 0.21 | 0.19 |

Standard errors are clustered at the LLM level (*p < 0.1, ** p < 0.05, ***p < 0.01). The sample includes non-manual employees with at least secondary education, drawn from the SHIW. The dependent variable is equal to 1 for public sector employees and to 0 for industrial sector employees. Grades are the final grades (normalized with respect to the maximum obtainable grade) obtained at the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: $C^1$ – i.e., reported crimes net of judicial efficiency – and $C^2$ – i.e., the principal component of *CPI*, *TRUST*, *GP* and $C^1$. Other controls include fixed effects for gender, cohort, and survey wave.

In Tables 6 and 7 we rely on SHIW data rather than on the LFS. Results should be interpreted with some caution given the relatively small number of observations and the large number of fixed effects that we include in the specification in order to control for the relevant unobserved heterogeneity. However, SHIW data allow us to use a second measure of ability, i.e., an index representing the grade obtained by individuals at their highest achieved educational level, which is available only for those with at least secondary education.

In Table 6 we consider years of schooling as the only ability measure and we restrict the analysis to non-manual employees with at least a secondary education diploma, for comparability with the sample for which we know final grade (see next paragraph). These results generally confirms previous ones, though the impact of education on the probability of joining the public sector is, on average, slightly weaker. Estimates for our parameter of interest, albeit only weakly significant, testify the presence of a detrimental effect of corruption, again concentrated among managers.

In Table 7 we focus the attention on ability as measured by school grades.[22] According to our findings, having obtained one additional grade-point, in a scale ranging from 0 (the lowest grade) to 10 (the highest grade), decreases the probability of joining the public sector (with respect to the manufacturing sector), but not for managers and other professionals. For this category, the impact of grades is differentiated across LLMs characterized by a different intensity of corruption, which negatively affects the propensity of more talented students to join the public sector. According to our estimates, one additional grade-point increases the likelihood of working for the public sector in a managerial position by 8 percentage points in low-corruption LLMs, and decreases it by 6 percentage points in high-corruption LLMs. Results are qualitatively similar if we use $C^2$ instead of $C^1$ to measure corruption.

### *4.2. The impact of corruption on workforce allocation*

In this section we inspect the impact of corruption on the effectiveness of the allocation of human resources. The latter is examined comparing individual abilities and the skill content of jobs to which workers are assigned. Mismatch may consist of either under-

[22] Grades are highly correlated with educational attainments, as those who obtain higher grades at secondary level are also those who are more likely to get tertiary education. Therefore, to avoid collinearity, we do not consider the two ability measures jointly interacted with corruption.

**Table 8**
Mismatch: under- and over-education.

| Dependent variable: | Under-education | | Over-education | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Public sector × $C^1$ | 0.045*** | 0.057*** | −0.000 | 0.003 |
| | (0.007) | (0.013) | (0.005) | (0.008) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 396,768 | 99,644 |
| Adj. R-squared | 0.04 | 0.06 | 0.04 | 0.03 |
| Public sector × $C^2$ | 0.022*** | 0.026*** | −0.001 | −0.003 |
| | (0.004) | (0.006) | (0.003) | (0.004) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 396,768 | 99,644 |
| Adj. R-squared | 0.04 | 0.06 | 0.04 | 0.03 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level and we consider two measures: $C^1$ – i.e., reported crimes net of judicial efficiency – and $C^2$ – i.e., the principal component of *CPI*, *TRUST*, *GP* and $C^1$.Other controls include fixed effects for gender, cohort, and survey wave.

education (an individual is assigned to a task which is on average undertaken by more educated workers) or over-education (an individual is assigned to a task which is on average undertaken by less educated workers).

As we have described before, under- and over-education are by and large as frequent in the public sector as they are elsewhere. The aim of our empirical strategy is once again to examine differential patterns between low- and high-corruption areas. Table 8 shows the results of the estimation of model (2). The coefficient associated to the interaction term between the public sector dummy and the measure of corruption is positive, suggesting that corruption increases the correlation between being in the public sector and the likelihood of being under-educated. These results hold for all employees and for the subset of those who stay at the top of the occupational hierarchy. On the other hand, we do not find any detectable effect in terms of over-education.

These patterns of skills mismatch might be, at least partially, a mechanical consequence of the negative selection phenomenon observed in the previous subsection. If corruption makes public employment relatively less attractive for the most educated, public agencies in corrupt areas will hire relatively less educated workers. Assuming that the tasks assigned to each agency do not vary with the level of corruption, under-education will arise as the obvious outcome of having to fill the same job positions with less educated employees. But under-education could also result from biased management practices, that may be more likely to occur where corruption is more intense. As an attempt to disentangle these two factors, in Table 9 we estimate a model identical to (2) but for two additional controls: the average skill content of professions present in each sector-LLM cell (measured as product of the nation-wide average of schooling in each profession and the share of professions in the cell) as well as the average education level in the same sector-LLM cell (measured with the average schooling of those employed in the cell).[23] These should respectively account for different educational endowments (and thus for the effects of selection) as well as for possible inflation in the number of high-level positions managed by public agencies.

The last consideration is also important to interpret the relationship between corruption levels and over-education. The absence of an effect for managers is expected: since the average skill content for managers typically corresponds to a university degree, very few cases (limited to individuals with post-tertiary studies) of over-education can exist for this professional category. On the other hand, over-education is not ex-ante unlikely for the whole population of non-manual workers. Column 3 of Table 9 indeed shows that, once average skill content and educational levels are accounted for, corruption is associated with an increase in over-education. The absence of effect in the main specification (column 3 of Table 8) can thus be attributed to not controlling for these factors. One possible interpretation is that inflation in the relative frequency of managers over all non-manual workers countervails the increased likelihood of highly educated workers to be assigned to non-managerial tasks.

As for under-education, we indeed find, as expected, that additional controls partly explain the distortionary effect of corruption: for example, under-education is more likely where the average schooling of employees is lower and where the average schooling required by the available job positions is higher. Nevertheless, our main effect still emerges, excluding that the estimates of Table 8 should be attributed uniquely to a mechanical consequence of bad workforce selection.

### *4.3. Robustness*

This section contains a first set of robustness checks. First, we examine whether our results hold when we modify the control group or use different sample selection rules. So far we have used the manufacturing sector as control group as it is relatively less

[23] This is the LLM-by-sector average of the average education level within professions based on the ISCO 3-digit classification.

**Table 9**
Under- and over-education: education and skill content corrections.

| Dependent variable: | Under-education | | Over-education | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Public sector × $C^1$ | 0.022*** | 0.031*** | 0.015*** | 0.011 |
| | (0.005) | (0.012) | (0.005) | (0.008) |
| Avg. skill content | 0.091*** | 0.075*** | −0.055*** | −0.032*** |
| | (0.003) | (0.009) | (0.003) | (0.005) |
| Avg. education level | −0.079*** | −0.076*** | 0.056*** | 0.026*** |
| | (0.002) | (0.006) | (0.002) | (0.004) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 396,768 | 99,644 |
| Adj. R-squared | 0.05 | 0.06 | 0.05 | 0.04 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Average skill content is the LLM-by-sector is measured as the product of the nation-wide average of schooling in each profession and the share of each professions in the sector-LLM cell; average educational level is the average schooling of the employed in the sector-LLM cell. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, and survey wave.

dependent on public spending and, therefore, arguably unaffected (or, at least, less affected) by corruption. In the first two columns of Table 10, however, we replicate our baseline results on selection, and under- and over-education extending the control group from manufacturing to the whole private sector.[24] Our main results are fully confirmed though the estimated effects are slightly smaller, thus implicitly suggesting that misallocation due to corruption somewhat extends to the private industries that interact more with the public sector.[25]

Moreover, having shown that corruption is more widespread in the South of Italy, we examine to what extent our results are driven by the traditional North-South divide and whether they still hold when we compare more homogeneous regions. The last two columns of Table 10 replicate our baseline results restricting the analysis to the LLMs located in the Center-North, obviously at the cost of losing a significant number of observations and territorial variability. The estimates of the coefficients associated with the interaction term in the selection equation (1) are fairly similar to those of our baseline specification. As far as misallocation is concerned, corruption continues to be significantly associated with under-education, especially among managers.

Second, we examine whether our patterns on the skill-biased impact of corruption on selection and allocation processes vary across different sections of the public sector. Therefore we replicate our baseline results distinguishing between public administration, education and healthcare. This is also useful to inspect the potential consequences of our sector-based definition of public sector (see section 2.1). Indeed, occupations in the administrative sector are almost by definition public sector jobs. In Italy, the majority of education-related jobs also lie under national or local government control. On the other hand, a more substantial minority of healthcare employees may actually be private sector workers whom we are incorrectly classifying as civil servants. This problem should, if anything, be more evident for non-managerial workers (since 'managers' in health-related sectors are mostly doctors, whose majority is – at least for part of their services – employed by national or local authorities).

Consistently with these considerations, in Table 11 we find that the association between schooling and the likelihood of having joined the public sector remains weaker in high-corruption areas for public administration and education, while the effects fades in the case of health services. In particular, for health-related jobs, a negative impact of corruption still remains for those in the most skilled professions, while the effect is zero if we also consider less skilled workers, for whom misclassification into the public sector is more likely. Our finding that, conditionally on being selected in the public sector, corruption increases the likelihood of being under-educated is fairly homogeneous across all employees and all areas of the public sector.

### *4.4. Endogeneity*

As discussed in subsection 3.3 we do not have (quasi-) experimental evidence to identify the causal impact of corruption on workforce selection and allocation in the public sector. The two main identification threats are the omitted variable bias and reverse causality. In what follows, we discuss the strategies we adopt to address these potential sources of endogeneity and to get closer to causal nexus.

[24] All results of this subsection are qualitatively similar if we use $C^2$ instead of $C^1$ as our measure of corruption.

[25] In Table A2 in the Appendix, we replicate the analysis using a continuous indicator of dependence on the public sector in place of the discrete indicator. More specifically, we map economic activities into the unit interval, capturing the dependence and/or the proximity between each economic sector of activity and the public sector, using the input-output matrix. The smallest values of this continuous measure correspond to sectors that do not interact with the public sector (e.g. the manufacturing sector); in contrast, larger values correspond to industries whose demand partly depends on public spending and/or that operate on regulated markets (e.g. electricity, water, waste disposal, construction, etc.). The results are qualitatively similar.

**Table 10**
Alternative samples.

| Dependent variable: | Employed in the public sector | | | |
|---|---|---|---|---|
| Sample: | **All sectors** | | **Center-North only** | |
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Years of schooling | 0.022*** | 0.046*** | 0.018*** | 0.030*** |
| | (0.001) | (0.001) | (0.001) | (0.002) |
| Years of schooling × $C^1$ | −0.002** | −0.009*** | −0.012*** | −0.017*** |
| | (0.001) | (0.002) | (0.004) | (0.006) |
| LLM × Prof. area FE | X | X | X | X |
| Observations | 753,043 | 135,127 | 262,935 | 58,978 |
| Adj. R-squared | 0.25 | 0.28 | 0.25 | 0.23 |
| Dependent variable: | Under-education | | | |
| Public sector × $C^1$ | 0.010*** | 0.024*** | 0.068*** | 0.097*** |
| | (0.004) | (0.007) | (0.015) | (0.028) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 753,048 | 135,127 | 262,939 | 58,978 |
| Adj. R-squared | 0.06 | 0.05 | 0.05 | 0.05 |
| Dependent variable: | Over-education | | | |
| Public sector × $C^1$ | 0.001 | −0.000 | 0.003 | 0.034* |
| | (0.003) | (0.003) | (0.012) | (0.019) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 753,048 | 135,127 | 262,939 | 58,978 |
| Adj. R-squared | 0.05 | 0.03 | 0.05 | 0.04 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. In the last two columns only workers residing in the Center-North are considered. The dependent variables are the following: 'employed in the public sector' is equal to 1 for public sector employees and to 0 for private sector (manufacturing sector) employees in the first (last) two columns; under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). Years of schooling are those corresponding to the highest educational attainment. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the private (manufacturing) sector in the first (last) two columns. Corruption is measured at the LLM level by $C^1$ – i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, and survey wave.

The first concern is related to potential omitted-variables bias. In the case of the selection phenomenon we study, where we believe these concerns may play a more relevant role, such biases may be induced by either individual or aggregate characteristics whose joint distribution with education changes at different levels of corruption, and that can affect education-based sorting into the public sector. Table 12 adds interactions between schooling and two observable individual characteristics, namely age and gender. While these factors have a significance of their own in terms of education-based selection, our coefficient of interest remains unchanged upon their inclusion.[26]

Aggregate variables may also be of relevance. For instance, in less developed areas the corruption might be more widespread and, at the same time, employment opportunities in the private sector might be lower. This would induce a spurious correlation between corruption and selection of the workforce towards the public sector. Our baseline specification includes LLM-fixed effects that should capture the impact of all potential omitted factors varying at that level, while exploiting the differential impact of corruption across individuals with a different level of schooling. However, our results might also be driven by other omitted variables correlated with corruption and implying differential effects similar to those produced by corruption itself. To address this point, we enrich the specification with other local controls aimed at capturing relevant economic dimensions that are both correlated with corruption and potentially liable to affect the link between individual human capital and occupational choices.[27] In the first column of Table 13 we include the (logarithm of the) value added per employee interacted with schooling as a determinant of selection into the public sector. The underlying idea is that better economic prospects at the local level (and any other variable correlated with economic development) might affect the education-based sorting between public and private sector. In the second column we include the unemployment rate at the LLM level: unemployment and corruption may be related through a number of channels and unemployment might affect the composition of the workforce willing to join the public sector. In the third column we include population density at the LLM level: this might affect both corruption and selection patterns, for instance because the scope of public administration can differ between urban and rural areas. In the fourth column we include the average firm size: one might expect that smaller firms are more opaque and at the same time the size of the private firm can shape the sorting of individuals between the private and the public sector. Finally, in the last column, we include the endowments of infrastructures that are correlated with

[26] In Table A3 in the Appendix, we also test the robustness of our results on misallocation and effort including age and gender interactions. In Table A4 we replicate our main results for specific sub-sectors of the public sector and on males and females separately.

[27] Estimates of a similarly extended model of misallocation are reported in Table A5 in the Appendix.

**Table 11**
Different sections of the public sector.

| Dependent variable: | Employed in the public sector | | | | | |
|---|---|---|---|---|---|---|
| Section of p.s.: | **Administration** | | **Education** | | **Healthcare** | |
| Professional area: | *All* | *Managers* | *All* | *Managers* | *All* | *Managers* |
| Schooling | 0.023*** | 0.022*** | 0.028*** | 0.035*** | 0.012*** | 0.056*** |
| | (0.001) | (0.002) | (0.001) | (0.002) | (0.001) | (0.002) |
| Schooling × $C^1$ | −0.009*** | −0.010*** | −0.008*** | −0.019*** | −0.000 | −0.005* |
| | (0.002) | (0.003) | (0.001) | (0.002) | (0.001) | (0.003) |
| LLM × Prof. area FE | X | X | X | X | X | X |
| Observations | 183,816 | 28,590 | 206,720 | 65,036 | 197,464 | 30,831 |
| Adj. R-squaredd | 0.35 | 0.34 | 0.51 | 0.31 | 0.30 | 0.39 |
| Dependent variable: | Under-education | | | | | |
| Section of p.s.: | **Administration** | | **Education** | | **Healthcare** | |
| Professional area: | *All* | *Managers* | *All* | *Managers* | *All* | *Managers* |
| Public sector × $C^1$ | 0.038*** | 0.038*** | 0.043*** | 0.066*** | 0.039*** | 0.034*** |
| | (0.007) | (0.014) | (0.008) | (0.014) | (0.008) | (0.012) |
| LLM FEs | X | X | X | X | X | X |
| Sector FEs | X | X | X | X | X | X |
| Observations | 183,840 | 28,590 | 206,735 | 65,036 | 197,488 | 30,831 |
| Adj. R-squared | 0.06 | 0.08 | 0.05 | 0.06 | 0.06 | 0.07 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: 'employed in the public sector' is equal to 1 for public sector employees and to 0 for private sector (manufacturing sector) employees in the first (last) two columns; under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). 'Schooling' indicates years of schooling corresponding to the highest educational attainment. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level by $C^1$, i.e. reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, and survey wave.

**Table 12**
Selection: age and gender interactions.

| Dependent variable: | Employed in the public sector | |
|---|---|---|
| Professional area: | *All* | *Managers* |
| Years of schooling | 0.035*** | 0.053*** |
| | (0.003) | (0.005) |
| Schooling × $C^1$ | −0.008*** | −0.013*** |
| | (0.001) | (0.001) |
| Schooling × Age | −0.001*** | −0.002*** |
| | (0.000) | (0.000) |
| Schooling × Female | −0.012*** | −0.021*** |
| | (0.001) | (0.002) |
| LLM × Prof. area FE | X | X |
| Observations | 396,764 | 99,644 |
| Adj. R-squared | 0.27 | 0.22 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. 'Schooling' indicates years of schooling corresponding to the highest educational attainment. Corruption is measured at the LLM level and we consider two measures: $C^1$ – i.e., reported crimes net of judicial efficiency – and $C^2$ – i.e., the principal component of *CPI*, *TRUST*, *GP* and $C^1$. Other controls include fixed effects for gender, cohort, and survey wave.

extent of corruption and at the same time can affect individual labor market choices increasing their geographical radius. Our main findings are robust to the inclusion of all these variables.

The second concern is related to reverse causality. Corruption may itself be the result of poorly selected public employees, while we are interested in the link *from* corruption *to* workforce selection and allocation. To address this issue we rely on an instrumental variable strategy. Our instruments are characterized by being pre-dated with respect to the hiring of current public employees. The first instrument measures past local dependence on public-sector demand. The underlying idea, implicitly supported by our previous findings, is that corruption episodes are more likely to occur where the role of public spending is more prominent. The second instrument measures the length of foreign domination spells. Having foreign rather than local rulers may have generated incentives for particular informal institutions, that may in turn be related to more widespread corruptive practices. Both instruments are computed at the province level.

As highlighted before, both our specifications include large sets of fixed effects. Thus exogeneity of our instruments with respect to the error term amounts to requiring the instrument not to be correlated with the variation of unobserved individual factors once LLM-fixed effects are partialled out. Thus, for instance, while areas with higher past dependence on the public sector or more protracted foreign domination may on average be characterized by systematically different patterns of workforce selection and/or

**Table 13**
Selection: adding aggregate controls.

| Dependent variable: | Employed in the public sector | | | | |
|---|---|---|---|---|---|
| Interacting variable ($Z$): | **V. add.** | **Unempl.** | **Density** | **Firm size** | **Infrastr.** |
| Professional area: | *All* | | | | |
| Schooling | 0.016*** | 0.016*** | 0.016*** | 0.016*** | 0.016*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Schooling × $C^1$ | −0.007*** | −0.003*** | −0.008*** | −0.004*** | −0.008*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Schooling × $Z$ | 0.002*** | −0.007*** | 0.000 | 0.006*** | −0.000 |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| LLM × Prof. area FE | X | X | X | X | X |
| Observations | 396,764 | 396,764 | 396,764 | 396,764 | 396,764 |
| Adj. R-squared | 0.27 | 0.27 | 0.27 | 0.27 | 0.27 |
| Professional area: | *Managers* | | | | |
| Schooling | 0.025*** | 0.026*** | 0.025*** | 0.026*** | 0.025*** |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Schooling × $C^1$ | −0.012*** | −0.006*** | −0.013*** | −0.005*** | −0.013*** |
| | (0.002) | (0.001) | (0.002) | (0.001) | (0.002) |
| Schooling × $Z$ | 0.005*** | −0.010*** | 0.002 | 0.011*** | 0.001 |
| | (0.001) | (0.001) | (0.003) | (0.001) | (0.001) |
| LLM × Prof. area FE | X | X | X | X | X |
| Observations | 99,644 | 99,644 | 99,644 | 99,644 | 99,644 |
| Adj. R-squared | 0.22 | 0.22 | 0.22 | 0.22 | 0.22 |

Standard errors are clustered at the LLM level (*p < 0.1, ** p < 0.05, ***p < 0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable is equal to 1 for public sector employees and to 0 for manufacturing sector employees. 'Schooling' indicates years of schooling corresponding to the highest educational attainment. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. We also include controls that might affect the public-private sector sorting between individuals with different educational level: the (logarithm of the) value added per employee (V. add.), population density (Density), unemployment rate (Unempl.), firm size, and availability of infrastructures (Infrastr.), all of which are measured at the LLM level – except for 'Infrastr.', measured at the province level – and standardized. Other controls include fixed effects for gender, cohort, and survey wave.

personnel allocation, this should not constitute a problem for our empirical strategy unless the variables used as instruments are correlated with individual characteristics (say, motivation or honesty) which may affect individual choice across areas and sectors.

IV estimates are reported in Table 14. Both instruments appear to be clearly related to corruption,[28] and the first stage F-statistic of the excluded instrument for the whole sample is well above the traditional threshold. The second-stage coefficients confirm the results of our baseline specification, thus attenuating the concerns of distortions attributable to reverse causality.

### *4.5. The impact of corruption on effort*

Unfortunately we do not possess data on individual performance or on the output produced by public employees. Therefore, we cannot evaluate whether the worsening of workforce selection and allocation processes due to corruption is also associated, in the aggregate, to a lower quality of the individual contribution to the local community. However, through the Labor Force Survey we can observe some individual inputs, as the number of hours worked. As shown in Table 1, public employees work on average fewer hours than private employees. The spread is around 10% larger, in absolute values, in areas with higher corruption levels than in those with lower corruption levels.

In Table 15 we estimate an equation similar to (2) with (the logarithm of) number of hours worked as dependent variable. We find that the coefficient associated to the interaction term between the public sector indicator and the measure of corruption is negative, suggesting that in areas with more corruption the number of hours worked by public employees relative to those in the manufacturing sector decreases. The coefficient for the subsample of managers and high-level professionals is again larger than that referring to all non-manual workers. We also report IV estimates, instrumenting corruption with the two instruments seen in the previous section, which confirm the previous results.

These results indicate that, besides worsening selection and allocation processes, corruption also leads to relatively lower effort by public employees. We argue that these public employees' characteristics and behavior can at least in part explain the negative relation between corruption levels and the quality of local public services, as documented by Fig. 3.

[28] In our specification, the coefficient of interest refers to the interaction $X \times C^1$, where $X$ is either years of schooling (selection models) or the indicator for employees working in the public sector (allocation and effort models). IV estimates are obtained by a 2SLS procedure, in which the first-stage equation is a regression of $X \times C^1$ on $Z \times C^1$, where $Z$ is the chosen instrument. Such regression includes the same controls included in the main equation. The relevant first-stage coefficient estimate is labeled '1st stage coeff.' in Tables 14 and 15. In the first panel of Table 14, we also report the results of regressions of $C^1$ on $Z$, run at the LLM level for the relevant sample, with no additional controls. The relevant coefficient estimate is labeled 'Pseudo 1st stage coeff.'. As this estimate does not vary with $X$, it is common across all IV analyses in the paper.

**Table 14**
IV estimates.

| Dependent variable: | Employed in the public sector | | | |
|---|---|---|---|---|
| Instrument: | **Past public dependence** | | **Foreign domination** | |
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Years of schooling | 0.016*** | 0.027*** | 0.016*** | 0.026*** |
| | (0.001) | (0.002) | (0.001) | (0.002) |
| Years of schooling × $C^1$ | −0.041*** | −0.071*** | −0.029*** | −0.034*** |
| | (0.007) | (0.012) | (0.006) | (0.009) |
| LLM × Prof. area FE | X | X | X | X |
| Observations | 396,764 | 99,644 | 396,764 | 99,644 |
| Adj. R-squared | 0.25 | 0.15 | 0.26 | 0.21 |
| Kleibergen-Paap F statistic | 41.88 | 35.01 | 30.57 | 27.34 |
| 1st stage coeff. | 0.384*** | 0.402*** | 0.057*** | 0.060*** |
| *Pseudo* 1st *stage coeff.* | 0.394*** | 0.410*** | 0.062*** | 0.067*** |
| Dependent variable: | Under-education | | | |
| Public sector × $C^1$ | 0.127*** | 0.083** | 0.078*** | 0.167*** |
| | (0.027) | (0.041) | (0.028) | (0.064) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 396,768 | 99,644 |
| Adj. R-squared | 0.04 | 0.06 | 0.04 | 0.05 |
| Kleibergen-Paap F statistic | 38.19 | 35.42 | 18.96 | 15.90 |
| Dependent variable: | Over-education | | | |
| Public sector × $C^1$ | −0.005 | −0.013 | −0.068** | 0.004 |
| | (0.016) | (0.027) | (0.027) | (0.039) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 396,768 | 99,644 |
| Adj. R-squared | 0.04 | 0.03 | 0.04 | 0.03 |
| Kleibergen-Paap F statistic | 38.19 | 35.42 | 18.96 | 15.90 |
| 1st stage coeff. | 0.329*** | 0.334*** | 0.038*** | 0.077*** |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: 'employed in the public sector' is equal to 1 for public sector employees and to 0 for manufacturing sector employees; under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). Years of schooling are those corresponding to the highest educational attainment of the individual. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level and we consider reported crimes net of judicial efficiency ($C^1$) instrumented with past public sector dependence (columns 1 and 2) and with the length of foreign domination spells (columns 3 and 4). Other controls include fixed effects for gender, cohort, and survey wave. '1st stage coeff.' is the coefficient of the first-stage regression, which is run using the same controls as the main specification. 'Pseudo 1st stage coefficient' reports the regression coefficient of $C^1$ on the instrument for the relevant sample.

These are not the only costs associated with corruption and the induced worst selection and allocation of the personnel in the public sector. Indeed, the existence of contestable rents might lead individuals to invest time, effort and resources to gain access to these rents, i.e. to invest these resources in unproductive and social costly activities.[29]

## 5. Conclusions

Our analysis highlights the distortionary effect of corruption on the patterns of selection and allocation of public sector employees. Because of the nature of the tasks assigned to and areas of activity spanned by public agencies, public employees are more educated with respect to their counterparts in the private sector. This gap is, however, thinner where corruption levels are higher. The education bias induced by corruption is particularly strong for professions at the top of the occupational hierarchy. Similar evidence is found if we consider further dimensions of human capital such as grades obtained at school. Besides affecting selection, corruption also contributes to deviating the education-based matching between workers and jobs: in areas with higher corruption levels, public employees are relatively more likely to be under-educated, that is to be assigned to tasks which are on average undertaken by more qualified personnel.

[29] See Aidt (2016) for a discussion on the intersection between the literature on corruption and that on rent-seeking in the Tullock sense.

**Table 15**
Labor supply and effort.

| Dependent variable: | Hours worked | | | | | |
|---|---|---|---|---|---|---|
| Instrument: | | | Past public dependence | | Foreign domination | |
| Professional area: | *All* | *Managers* | *All* | *Managers* | *All* | *Managers* |
| Public sector × $C^1$ | −0.017*** | −0.030*** | −0.060*** | −0.130*** | −0.090*** | −0.224*** |
| | (0.004) | (0.010) | (0.015) | (0.038) | (0.026) | (0.069) |
| LLM FEs | X | X | X | X | X | X |
| Sector FEs | X | X | X | X | X | X |
| Observations | 351,347 | 86,931 | 351,347 | 86,931 | 351,347 | 86,931 |
| Adj. R-squared | 0.40 | 0.42 | 0.40 | 0.42 | 0.40 | 0.42 |
| Kleibergen-Paap F statistic | | | 38.69 | 35.55 | 18.82 | 15.71 |
| 1st stage coeff. | | | 0.330*** | 0.338*** | 0.038*** | 0.034*** |

Standard errors are clustered at the LLM level (*p < 0.1, ** p < 0.05, ***p < 0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable is (the logarithm of) hours worked. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. In columns 3 and 4, corruption is instrumented with past public sector dependence; in columns 5 and 6, corruption is instrumented with the length of foreign domination spells. Other controls include fixed effects for gender, cohort, survey wave, and an indicator for whether the individual works part-time. '1st stage coeff.' is the coefficient of the first-stage regression, which is run using the same controls as the main specification.

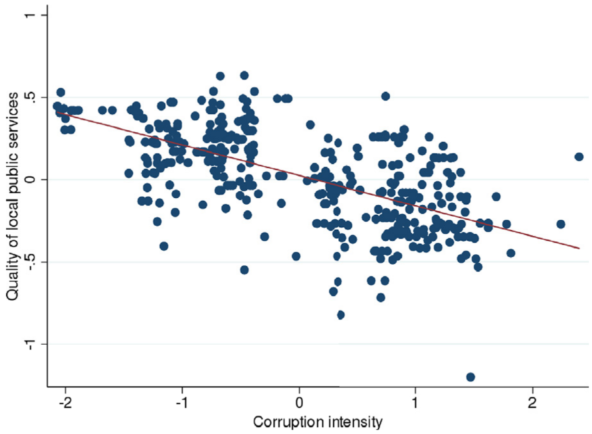


**Fig. 3.** Corruption intensity ($C^1$) and quality of local public services (from Camussi and Mancini, 2016) across LLMs.

Although these results cannot be interpreted in causal terms, since we do not have an identification strategy that unequivocally demonstrates a causal link, they do not seem to be driven by confounding factors at the local level and by reverse causation.

The comparative analysis of our results thus suggests that higher levels of corruption are associated with a poorer ability of the public sector to select and allocate workers. Hence – if one believes that the workforce's human capital is conducive toward better decision making – where corruption is high, public administrations will tend to adopt socially inefficient decisions and to remunerate individual less in terms of schooling ability than of other (unobserved) ability traits such as soft skills, relational capital or craftiness.

The eradication of corruption or, at least, the dampening of its implications has long been a major objective of government effort. Actions taken by authorities usually rest on ex-post, repressive measures, which are sometimes accompanied by ex-ante, preventive provisions. The latter often take the form of a requirement for individual agencies to implement "in-house" anti-corruptive programs under governmental supervision. In light of the evidence presented in this paper, one may suspect that the administrations' ability to exert anti-corruptive self-monitoring might be hindered by corruption itself. Indeed, existing levels of crime in the environment may have contributed to the selection of a workforce which will, in general, be more likely to be misallocated as well as less prone to take up action against corruption if called to do so. The risk is that self-regulation aimed at overcoming corruption may work well only where corruption is already rare, and fare rather poorly where it is more intense. Hence our results suggest caution against over-estimating the additional benefits of ex-ante, decentralized provisions.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ejpoleco.2019.07.007.

## Appendix

**Table A1**
Clustering at the Province level.

| Dependent variable: | Employed in the public sector | | Under-education | | Hours worked | |
|---|---|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* | *All* | *Managers* |
| Years of schooling | 0.016*** | 0.025*** | | | | |
| | (0.001) | (0.002) | | | | |
| Years of schooling × $C^1$ | −0.008*** | −0.013*** | | | | |
| | (0.001) | (0.002) | | | | |
| Public sector × $C^1$ | | | 0.045*** | 0.057*** | −0.017*** | −0.030** |
| | | | (0.009) | (0.013) | (0.004) | (0.011) |
| LLM × Prof. area FE | X | X | | | | |
| LLM FEs | | | X | X | X | X |
| Sector FEs | | | X | X | X | X |
| Observations | 396,764 | 99,644 | 396,768 | 99,644 | 351,347 | 86,931 |
| Adj. R-squared | 0.27 | 0.22 | 0.04 | 0.06 | 0.40 | 0.42 |

Standard errors are clustered at the Province level ($^*$p < 0.1, $^{**}$ p < 0.05, $^{***}$p < 0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: 'employed in the public sector' is equal to 1 for public sector employees and to 0 for manufacturing sector employees; under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification); hours worked are log-transformed. Years of schooling are those corresponding to the highest educational attainment of the individual. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, survey wave, and an indicator for whether the individual works part-time in columns 5 and 6.

**Table A2**
Under- and over-education: exposure.

| Dependent variable: | Under-education | | Over-education | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Exposure × $C^1$ | 0.012*** | 0.026*** | 0.001 | 0.002 |
| | (0.004) | (0.009) | (0.003) | (0.004) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 753,048 | 135,127 | 753,048 | 135,127 |
| Adj. R-squared | 0.06 | 0.05 | 0.05 | 0.03 |

Standard errors are clustered at the LLM level ($^*$p < 0.1, $^{**}$ p < 0.05, $^{***}$p < 0.01). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education (over-education) is equal to 1 if the employee has a number of years of schooling below the 25th percentile (above the 75th percentile) of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). 'Exposure' is a continuous variable measuring the exposure of the worker's sector to the public sector. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, and survey wave.

**Table A3**
Age and gender interactions.

| Dependent variable: | Under-education | | Hours worked | |
|---|---|---|---|---|
| Professional area: | *All* | *Managers* | *All* | *Managers* |
| Public sector × $C^1$ | 0.045*** | 0.056*** | −0.017*** | −0.030*** |
| | (0.007) | (0.013) | (0.004) | (0.010) |
| Public sector × Age | −0.017*** | −0.015*** | 0.002** | −0.005** |
| | (0.001) | (0.003) | (0.001) | (0.002) |
| Public sector × Female | −0.038*** | −0.057*** | −0.004 | −0.031*** |
| | (0.007) | (0.012) | (0.004) | (0.010) |
| LLM FEs | X | X | X | X |
| Sector FEs | X | X | X | X |
| Observations | 396,768 | 99,644 | 351,347 | 86,931 |
| Adj. R-squared | 0.05 | 0.06 | 0.40 | 0.42 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification); hours worked are log-transformed. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, survey wave, and an indicator for whether the individual works part-time in columns 3 and 4.

**Table A4**
Gender heterogeneity.

| Dependent variable: | Employed in the public sector | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Section of public sector: | All | | Administration | | Education | | Healthcare | |
| Gender: | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* | *Male* | *Female* |
| Schooling | 0.025*** | 0.008*** | 0.019*** | 0.031*** | 0.029*** | 0.027*** | 0.028*** | −0.002* |
| | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) |
| Schooling × $C^1$ | −0.011*** | −0.002*** | −0.008*** | −0.009*** | −0.004* | −0.009*** | −0.004* | 0.002 |
| | (0.002) | (0.001) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) | (0.002) |
| LLM × Prof. area FE | X | X | X | X | X | X | X | X |
| Observations | 163,515 | 233,228 | 110,139 | 73,601 | 85,574 | 121,080 | 87,474 | 109,933 |
| Adj. R-squared | 0.33 | 0.19 | 0.38 | 0.37 | 0.47 | 0.45 | 0.31 | 0.19 |
| Dependent variable: | Under-education | | | | | | | |
| Public sector × $C^1$ | 0.054*** | 0.037*** | 0.050*** | 0.018* | 0.047*** | 0.035*** | 0.041*** | 0.036*** |
| | (0.009) | (0.008) | (0.009) | (0.009) | (0.011) | (0.008) | (0.010) | (0.010) |
| LLM FEs | X | X | X | X | X | X | X | X |
| Sector FEs | X | X | X | X | X | X | X | X |
| Observations | 163,526 | 233,238 | 110,172 | 73,658 | 85,620 | 121,105 | 87,507 | 109,970 |
| Adj. R-squared | 0.05 | 0.05 | 0.05 | 0.09 | 0.05 | 0.06 | 0.05 | 0.08 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variables are the following: 'employed in the public sector' is equal to 1 for public sector employees and to 0 for manufacturing sector employees; under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). 'Schooling' are years of schooling corresponding to the highest educational attainment of the individual. 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Other controls include fixed effects for gender, cohort, and survey wave.

**Table A5**
Under-education: adding aggregate controls.

| Dependent variable: | Under-education | | | | |
|---|---|---|---|---|---|
| Interacting variable ($Z$): | **V. add.** | **Unempl.** | **Density** | **Firm size** | **Infrastr.** |
| Professional area: | *All* | | | | |
| Public sector × $C^1$ | 0.045*** | 0.026*** | 0.044*** | 0.036*** | 0.044*** |
| | (0.007) | (0.008) | (0.007) | (0.007) | (0.007) |
| Public sector × $Z$ | 0.001 | 0.021*** | 0 .007*** | −0.013*** | 0.006* |
| | (0.003) | (0.003) | (0.002) | (0.004) | (0.003) |
| LLM × Prof. area FE | X | X | X | X | X |
| Observations | 396,768 | 396,768 | 396,768 | 396,768 | 396,768 |
| Adj. R-squared | 0.04 | 0.04 | 0.04 | 0.04 | 0.04 |
| Professional area: | *Managers* | | | | |
| Public sector × $C^1$ | 0.057*** | 0.042*** | 0.055*** | 0.050*** | 0.057*** |
| | (0.013) | (0.014) | (0.013) | (0.015) | (0.013) |
| Public sector × $Z$ | −0.001 | 0.015** | 0.007** | −0.010 | 0.001 |
| | (0.006) | (0.006) | (0.003) | (0.008) | (0.005) |
| LLM × Prof. area FE | X | X | X | X | X |
| Observations | 99,644 | 99,644 | 99,644 | 99,644 | 99,644 |
| Adj. R-squared | 0.06 | 0.06 | 0.06 | 0.06 | 0.06 |

Standard errors are clustered at the LLM level (*$p < 0.1$, ** $p < 0.05$, ***$p < 0.01$). The sample includes non-manual employees, drawn from the LFS. The dependent variable under-education is equal to 1 if the employee has a number of years of schooling below the 25th percentile of the years of schooling distribution of the jobs he/she is assigned to (based on the ISCO 3-digit classification). 'Public sector' indicates whether a worker is employed by the public sector or, alternatively, by the manufacturing sector. Corruption is measured at the LLM level by $C^1$, i.e., reported crimes net of judicial efficiency. Further controls are included: the (logarithm of the) value added per employee (V. add.), population density (Density), unemployment rate (Unempl.), firm size, and availability of infrastructures (Infrastr.), all of which are measured at the LLM level – except for 'Infrastr.', measured at the province level – and standardized. Other controls include fixed effects for gender, cohort, and survey wave.

## References


Acemoglu, D., Robinson, J.A., 2012. Why Nations Fail: the Origins of Power, Prosperity, and Poverty. Random House, New York.
Acemoglu, D., Verdier, T., 1998. Property rights, corruption and the allocation of talent: a general equilibrium approach. Econ. J. 108 (23), 186–202.
Aidt, T.S., 2016. Rent seeking and the economics of corruption. Consititut. Pol. Econ. 27 (2), 142–157.
Banerjee, R., Baul, T., Rosenblat, T., 2015. On self selection of the corrupt into the public sector. Econ. Lett. 127, 43–46.
Barfort, S., Harmon, N., Hjorth, F., Olsen, A.L., 2017. Sustaining Honesty in Public Service: the Role of Selection. Working paper.
Barone, G., Mocetti, S., 2011. Tax morale and public spending inefficiency. Int. Tax Publ. Financ. 18 (6), 724–749.
Barone, G., Mocetti, S., 2014. Natural disasters, growth and institutions: a tale of two earthquakes. J. Urban Econ. 84, 52–66.
Brandolini, A., Cannari, L., 1994. Methodological appendix: the Bank of Italy Survey of household income and wealth. In: Ando, A., Guiso, L., Visco, I. (Eds.), Saving and the Accumulation of Wealth. Essays on Italian Households and Government Saving Behavior. Cambridge University Press.
Brollo, F., Nannicini, T., Perotti, R., Tabellini, G., 2013. The political resource curse. Am. Econ. Rev. 103 (5), 1759–1796.
Camussi, S., Mancini, A.L., 2016. Individual Trust: Does Quality of Public Services Matter? Bank of Italy. Working Paper no. 1069.
Charron, N., Lapuente, V., Dijkstra, L., 2014. Regional governance matters: quality of government within European Union member states. Reg. Stud. 48 (1), 68–90.
Clausen, B., Kraay, A., Nyiri, Z., 2011. Corruption and confidence in public institutions: evidence from a global survey. World Bank Econ. Rev. 25 (2), 212–249.
Colonnelli, E., Prem, M., Teso, E., 2018. Patronage and Selection in Public Sector Organizations. mimeo.
Correia, S., 2017. Linear Models with High-Dimensional Fixed Effects: an Efficient and Feasible Estimator. Working Paper.
Cowley, E., Smith, S., 2014. Motivation and mission in the public sector: evidence from the World Values Survey. Theory Decis. 76 (2), 241–263.
Del Monte, A., Papagni, E., 2001. Public expenditure, corruption, and economic growth: the case of Italy. Eur. J. Political Econ. 17 (1), 1–16.
Del Monte, A., Papagni, E., 2007. The determinants of corruption in Italy: regional panel data analysis. Eur. J. Political Econ. 23 (2), 379–396.
Di Liberto, A., Sideri, M., 2015. Past dominations, current institutions and the Italian regional economic performance. Eur. J. Political Econ. 38, 12–41.
Ferraz, C., Finan, F., 2011. Electoral accountability and corruption: evidence from the audits of local governments. Am. Econ. Rev. 101 (4), 1274–1311.
Fisman, R., Gatti, R., 2012. Decentralization and corruption: evidence across countries. J. Public Econ. 83 (3), 325–345.
Fisman, R., Miguel, E., 2007. Corruption, norms, and legal enforcement: evidence from diplomatic parking tickets. J. Public Econ. 115 (6), 1020–1048.
Giorgiantonio, C., Orlando, T., Palumbo, G., Rizzica, L., 2016. Incentives and Selection in Public Employment. Bank of Italy. Occasional Paper no. 342.
Golden, M., Picci, L., 2005. Proposal for a new measure of corruption, illustrated with Italian data. Econ. Pol. 17 (1), 37–75.
Hanna, R., Wang, S.Y., 2017. Dishonesty and selection into public service: evidence from India. Am. Econ. J. Econ. Policy 9 (3), 262–290.
Hessami, Z., 2014. Political corruption, public procurement, and budget composition: theory and evidence from OECD countries. Eur. J. Political Econ. 34, 372–389.
Knack, S., Keefer, P., 1995. Institutions and economic performance: cross-country tests using alternative institutional measures. Econ. Pol. 7 (3), 207–227.
La Porta, R., Lopez de Silanes, F., Shleifer, A., Vishny, R., 1999. The quality of government. J. Law Econ. Organ. 15 (1), 222–279.
Lochner, L., 2004. Education, work, and crime: a human capital approach. Int. Econ. Rev. 45 (3), 811–843.
Madrian, B., 1994. Employment-based health insurance and job mobility: is there evidence of job-lock? Q. J. Econ. 109 (1), 27–54.
Mauro, P., 1995. Corruption and growth. Q. J. Econ. 110 (3), 681–712.
Mauro, P., 1998. Corruption and the composition of government expenditure. J. Public Econ. 69 (2), 263–279.
Mo, P.H., 2001. Corruption and economic growth. J. Comp. Econ. 29 (1), 66–79.
Moulton, B., 1990. An illustration of a pitfall in estimating the effects of aggregate variables on micro units. Rev. Econ. Stat. 72 (2), 334–338.
Murphy, K., Shleifer, A., Vishny, R.W., 1991. The allocation of talent: implications for growth. Q. J. Econ. 106 (2), 503–530.
Olken, B.A., 2009. Corruption perceptions vs. corruption reality. J. Public Econ. 93 (7), 950–964.

Olken, B.A., Barron, P., 2009. The simple economics of extortion: evidence from trucking in Aceh. J. Political Econ. 117 (3), 417–452.
Oreopoulos, P., Salvanes, K.G., 2011. Priceless: the nonpecuniary benefits of schooling. J. Econ. Perspect. 25 (1), 159–184.
Petrarca, I., Ricciuti, R., 2013. The historical economics of corruption and development within Italy. Int. J. Monet. Econ. Financ. 6, 52–66.
Rajan, G.R., Zingales, L., 1998. Financial dependence and growth. Am. Econ. Rev. 88 (3), 559–586.
Rizzica, L., 2016. Why Go Public? A Study of the Individual Determinants of Public Sector Employment Choice. Bank of Italy. Occasional Paper no. 343.
Rizzica, L., Tonello, M., 2015. Exposure to Media and Corruption Perceptions. Bank of Italy. Working Paper no. 1043.
Shleifer, A., Vishny, R.W., 1993. Corruption. Q. J. Econ. 108 (3), 599–617.
Uslaner, E.M., 2004. Trust and corruption. In: Lambsdorf, J.G., Taube, M., Schramm, M. (Eds.), The New Institutional Economics of Corruption. Routledge, London.